

The McMillan theorem for colored branching processes and dimensions of random fractals

Victor I. Bakhtin

bakhtin@tut.by

For simplest colored branching processes we prove an analog to the McMillan theorem and calculate Hausdorff dimensions of random fractals defined in terms of the limit behavior of empirical measures generated by finite genetic lines. In this setting the role of Shannon's entropy is played by the Kullback–Leibler divergence and the Hausdorff dimensions are computed by means of the so-called Billingsley–Kullback entropy, defined in the paper.

Keywords: *random fractal, Hausdorff dimension, colored branching process, basin of the empirical measure, spectral potential, Billingsley–Kullback entropy, Kullback action, maximal dimension principle*

2010 Mathematics Subject Classification: 28A80, 37F35, 60J80

Let us consider the finite set $X = \{1, \dots, r\}$, whose elements denote different colors, and a vector $(\mu(1), \dots, \mu(r)) \in [0, 1]$. A simplest colored branching process can be defined as an evolution of a population in which all individuals live the same fixed time and then, when the lifetime ends, each individual generates (independently of others) a random set of “children” containing individuals of colors $1, \dots, r$ with probabilities $\mu(1), \dots, \mu(r)$ respectively. We will suppose that the evolution starts with a unique initial individual. It is suitable to represent this process as a random genealogical tree with individuals as vertices and each vertex connected by edges with its children. Denote by X_n the set of all genetic lines of length n (that survive up to generation n). The colored branching process can degenerate (when it turns out that starting from some n all the sets X_n are empty) or, otherwise, evolve endlessly. Every genetic line $x = (x_1, \dots, x_n) \in X_n$ generates an empirical measure $\delta_{x,n}$ on the set of colors X by the following rule: for each $i \in X$ the value of $\delta_{x,n}(i)$ is the fraction of those coordinates of the vector (x_1, \dots, x_n) that coincide with i .

Let ν be an arbitrary probability measure on X . The analog to the McMillan theorem that will be proved below asserts that under condition of nondegeneracy of the colored branching process the cardinality of the set $\{x \in X_n \mid \delta_{x,n} \approx \nu\}$ has an almost sure asymptotics of order $e^{-n\rho(\nu, \mu)}$, where

$$\rho(\nu, \mu) = \sum_{i \in X} \nu(i) \ln \frac{\nu(i)}{\mu(i)}.$$

Formally, the value of $\rho(\nu, \mu)$ coincides with the usual Kullback–Leibler divergence and differs from the latter only in the fact that in our setting the measure μ is not probability and so $\rho(\nu, \mu)$ can be negative.

In the paper we investigate also random fractals defined in terms of the sequence of empirical measures $\delta_{x,n}$ limit behavior. Let X_∞ be the set of infinite genetic lines. Fix an arbitrary vector $\theta = (\theta(1), \dots, \theta(r)) \in (0, 1)^r$ and define the following metrics on X_∞ :

$$\text{dist}(x, y) = \prod_{t=1}^n \theta(x_t), \quad \text{where } n = \inf \{t \mid x_t \neq y_t\} - 1.$$

Denote by V any set of probability measures on X . It will be proved, in particular, that under the condition of nondegeneracy of the colored branching process

$$\dim_H \{x \in X_\infty \mid \delta_{x,n} \rightarrow V\} = \sup_{\nu \in V} d(\nu, \mu, \theta)$$

almost surely, where $d(\nu, \mu, \theta)$ is the *Billingsley–Kullback entropy* defined below.

The paper can be divided into two parts. The first one (sections 1–5) contains known results; some of them have been modified in a certain way for the convenience of use in what follows. Anyway, most of them are proved below for the completeness and convenience of the reader. The second part (sections 6–9) contains new results.

In addition, we note that all the results of the paper can be easily extended to Moran’s self-similar geometric constructions in \mathbb{R}^n , but we will not do that.

1 The spectral potential

Let X be an arbitrary finite set. Denote by $B(X)$ the space of all real-valued functions on X , by $M(X)$ the set of all positive measures on X , and by $M_1(X)$ the collection of all probability distributions on X .

Every measure $\mu \in M(X)$ determines a linear functional on $B(X)$ of the form

$$\mu[f] = \int_X f d\mu = \sum_{x \in X} f(x) \mu(x).$$

It is easily seen that this functional is *positive* (i. e., takes nonnegative values on non-negative functions). If, in addition, the measure μ is probability then this functional is normalized (takes the value 1 on the unit function).

Consider the nonlinear functional

$$\lambda(\varphi, \mu) = \ln \mu[e^\varphi], \tag{1.1}$$

where $\varphi \in B(X)$ and $\mu \in M(X)$. We will call it the *spectral potential*. Evidently, it is monotone (if $\varphi \geq \psi$ then $\lambda(\varphi, \mu) \geq \lambda(\psi, \mu)$), additively homogeneous (that is, $\lambda(\varphi + t, \mu) = \lambda(\varphi, \mu) + t$ for each constant t), and analytic in φ .

Define a family of probability measures μ_φ on X , depending on the functional parameter $\varphi \in B(X)$, by means of the formula

$$\mu_\varphi[f] = \frac{\mu[e^\varphi f]}{\mu[e^\varphi]}, \quad f \in B(X).$$

Evidently, each measure μ_φ is equivalent to μ and has the density $e^{\varphi - \lambda(\varphi, \mu)}$ with respect to μ .

Let us compute the first two derivatives of the spectral potential with respect to the argument φ . Introduce the notation

$$\lambda'(\varphi, \mu)[f] = \left. \frac{d\lambda(\varphi + tf, \mu)}{dt} \right|_{t=0}.$$

This is nothing more than the derivative of the spectral potential in the direction f at the point φ . An elementary computation shows that

$$\lambda'(\varphi, \mu)[f] = \left. \frac{d \ln \mu[e^{\varphi+tf}]}{dt} \right|_{t=0} = \frac{\mu[e^{\varphi}f]}{\mu[e^{\varphi}]} = \mu_{\varphi}[f]. \quad (1.2)$$

In other words, the derivative $\lambda'(\varphi, \mu)$ coincides with the probability measure μ_{φ} . Then put

$$\lambda''(\varphi, \mu)[f, g] = \left. \frac{\partial^2 \lambda(\varphi + tf + sg, \mu)}{\partial s \partial t} \right|_{s,t=0}$$

and compute this derivative using just obtained formula (1.2):

$$\lambda''(\varphi, \mu)[f, g] = \left. \frac{\partial}{\partial s} \left(\frac{\mu[e^{\varphi+sg}f]}{\mu[e^{\varphi+sg}]} \right) \right|_{s=0} = \frac{\mu[e^{\varphi}fg]}{\mu[e^{\varphi}]} - \frac{\mu[e^{\varphi}f]\mu[e^{\varphi}g]}{(\mu[e^{\varphi}])^2} = \mu_{\varphi}[fg] - \mu_{\varphi}[f]\mu_{\varphi}[g].$$

In probability theory the expression $\mu_{\varphi}[t]$ is usually called the expectation of the random variable f with respect to the probability distribution μ_{φ} , and the expression $\mu_{\varphi}[fg] - \mu_{\varphi}[f]\mu_{\varphi}[g]$ is called the covariance of random variables f and g . In particular, the second derivative

$$\left. \frac{d^2 \lambda(\varphi + tf, \mu)}{dt^2} \right|_{t=0} = \mu_{\varphi}[f^2] - \mu_{\varphi}[f]^2 = \mu_{\varphi}[(f - \mu_{\varphi}[f])^2]$$

is equal to the variance of the random variable f with respect to the distribution μ_{φ} . Since the variance is nonnegative it follows that the spectral potential is convex in φ .

2 The Kullback action

Denote by $B^*(X)$ the space of all linear functionals on $B(X)$. Then, obviously,

$$M_1(X) \subset M(X) \subset B^*(X).$$

The following functional of two arguments $\nu \in B^*(X)$ and $\mu \in M(X)$ will be called the *Kullback action*:

$$\rho(\nu, \mu) = \begin{cases} \mu[\varphi \ln \varphi] = \nu[\ln \varphi], & \text{if } \nu \in M_1(X) \text{ and } \nu = \varphi\mu, \\ +\infty & \text{in all other cases.} \end{cases} \quad (2.1)$$

To be more precise, the “all other cases” fit into at least one of the three categories: a) singular w. r. t. μ probability measures ν , b) nonnormalized functionals ν , and c) nonpositive functionals ν .

In the literature, as far as I know, this functional have been defined only for probability measures ν and μ . Different authors call it differently: the relative entropy, the

deviation function, the Kullback–Leibler information function, the Kullback–Leibler divergence.

When ν is a probability measure the Kullback action can be defined by the explicit formula

$$\rho(\nu, \mu) = \sum_{x \in X} \nu(x) \ln \frac{\nu(x)}{\mu(x)}. \quad (2.2)$$

In particular, if $\mu(x) \equiv 1$ then the Kullback action differs only in sign from Shannon's entropy

$$H(\nu) = - \sum_{x \in X} \nu(x) \ln \nu(x). \quad (2.3)$$

In the case of probability measure μ the Kullback action is nonnegative and vanishes only if $\nu = \mu$. Indeed, if the functional ν is not an absolutely continuous with respect to μ probability measure then $\rho(\nu, \mu) = +\infty$. Otherwise, if ν is a probability measure of the form $\nu = \varphi\mu$ then from Jensen's inequality and strong convexity of the function $f(x) = x \ln x$ it follows that

$$\rho(\nu, \mu) = \mu[f(\varphi)] \geq f(\mu[\varphi]) = 0$$

(so long as $\mu[\varphi] = \nu[1] = 1$), and the equality $\rho(\nu, \mu) = 0$ holds if and only if φ is constant almost everywhere and, respectively, ν coincides with μ .

Every measure $\mu \in M(X)$ can be put down in the form $\mu = c\mu_1$, where $c = \mu[1]$ and $\mu_1 \in M_1(X)$. If $\nu \in M_1(X)$ then (2.2) implies

$$\rho(\nu, \mu) = \rho(\nu, \mu_1) - \ln c \geq -\ln \mu[1]. \quad (2.4)$$

In case $\nu \notin M_1(X)$ this inequality holds all the more since the Kullback action is infinite.

Theorem 2.1 *The spectral potential and the Kullback action satisfy the Young inequality*

$$\rho(\nu, \mu) \geq \nu[\psi] - \lambda(\psi, \mu), \quad (2.5)$$

that turns into equality if and only if $\nu = \mu_\psi$.

Proof. If $\rho(\nu, \mu) = +\infty$ then the Young inequality is trivial. If $\rho(\nu, \mu) < +\infty$ then by the definition of Kullback action the functional ν is an absolutely continuous probability measure of the form $\nu = \varphi\mu$, where φ is a nonnegative density. In this case

$$\lambda(\psi, \mu) = \ln \mu[e^\psi] \geq \ln \int_{\varphi > 0} e^\psi d\mu = \ln \int_{\varphi > 0} e^{\psi - \ln \varphi} d\nu = \ln \nu[e^{\psi - \ln \varphi}] \geq \nu[\psi - \ln \varphi]$$

(at the last step we have used Jensen's inequality and concavity of the logarithm function). Since $\rho(\nu, \mu) = \nu[\ln \varphi]$, this formula implies inequality (2.5).

Recall that $\mu_\psi = e^{\psi - \lambda(\psi, \mu)}\mu$. So if $\nu = \mu_\psi$ then by definition

$$\rho(\nu, \mu) = \nu[\psi - \lambda(\psi, \mu)] = \nu[\psi] - \lambda(\psi, \mu).$$

Vice versa, assume that $\rho(\nu, \mu) = \nu[\psi] - \lambda(\psi, \mu)$. Then subtract from the above equality the Young inequality $\rho(\nu, \mu) \geq \nu[\varphi] - \lambda(\varphi, \mu)$. We obtain

$$\lambda(\varphi, \mu) - \lambda(\psi, \mu) \geq \nu[\varphi - \psi].$$

From this follows that $\nu = \lambda'(\psi, \mu)$. Finally, $\lambda'(\psi, \mu)$ coincides with μ_ψ . \square

Theorem 2.2 *The Kullback action $\rho(\nu, \mu)$ is the Legendre transform w. r. t. ν of the spectral potential:*

$$\rho(\nu, \mu) = \sup_{\psi \in B(X)} \{ \nu[\psi] - \lambda(\psi, \mu) \}, \quad \nu \in B^*(X), \quad \mu \in M(X). \quad (2.6)$$

Proof. By the Young inequality the left hand side of (2.6) is not less than the right one. Therefore it is enough to associate with any functional $\nu \in B^*(X)$ a family of functions ψ_t , depending on the real-valued parameter t , on which the equality in (2.6) is attained.

At first, suppose that ν is an absolutely continuous with respect to μ probability measure of the form $\nu = \varphi\mu$, where φ is a nonnegative density. Consider the family of functions

$$\psi_t(x) = \begin{cases} \ln \varphi(x), & \varphi(x) > 0, \\ -t, & \varphi(x) = 0. \end{cases}$$

When $t \rightarrow +\infty$ we have the following relations

$$\begin{aligned} \mu[e^{\psi_t}] &= \int_{\varphi>0} \varphi d\mu + \int_{\varphi=0} e^{-t} d\mu \longrightarrow \int_X \varphi d\mu = 1, \\ \nu[\psi_t] &= \int_{\varphi>0} \varphi \ln \varphi d\mu + \int_{\varphi=0} -t\varphi d\mu = \mu[\varphi \ln \varphi], \end{aligned}$$

$$\nu[\psi_t] - \lambda(\psi_t, \mu) = \nu[\psi_t] - \ln \mu[e^{\psi_t}] \longrightarrow \mu[\varphi \ln \varphi] = \rho(\nu, \mu),$$

and so (2.6) is proved.

In all the other cases, when ν is not an absolutely continuous probability measure, by definition $\rho(\nu, \mu) = +\infty$. Let us examine this cases one after another.

If ν is a singular relative to μ probability measure, then there exists $x_0 \in X$ such that $\mu(x_0) = 0$ and $\nu(x_0) > 0$. In this case consider the family of functions

$$\psi_t(x) = \begin{cases} t, & x = x_0, \\ 0, & x \neq x_0. \end{cases}$$

It is easily seen that

$$\nu[\psi_t] - \lambda(\psi_t, \mu) \geq t\nu(x_0) - \ln \mu[e^{\psi_t}] \geq t\nu(x_0) - \ln \mu[1].$$

The right hand side of the above formula goes to $+\infty$ while t increases and (2.6) holds again.

If the functional ν is not normalized then put $\psi_t = t$. Then the expression

$$\nu[\psi_t] - \lambda(\psi_t, \mu) = \nu[t] - \ln \mu[e^t] = t(\nu[1] - 1) - \ln \mu[1]$$

is unbounded from the above and hence (2.6) is still valid.

Finally, if the functional ν is not positive then there exists a nonnegative function φ such that $\nu[\varphi] < 0$. Consider the family $\psi_t = -t\varphi$, where $t > 0$. For it

$$\nu[\psi_t] - \lambda(\psi_t, \mu) \geq -t\nu[\varphi] - \lambda(0, \mu) \longrightarrow +\infty$$

as $t \rightarrow +\infty$, and (2.6) remains in force. \square

Corollary 2.3 *The functional $\rho(\cdot, \mu)$ is convex and lower semicontinuous on $B^*(X)$.*

Proof. These are properties of the Legendre transform. \square

3 The local large deviations principle and the McMillan theorem

As above, we keep to the following notation: X is a finite set, $B(X)$ stands for the space of real-valued functions on X , $B^*(X)$ is the space of linear functionals on $B(X)$, $M_1(X)$ is the set of all probability measures on X , and $M(X)$ is the set of all positive measures on X .

To each finite sequence $x = (x_1, \dots, x_n) \in X^n$ let us correspond an *empirical measure* $\delta_{x,n} \in M_1(X)$ which is supported on the set $\{x_1, \dots, x_n\}$ and assigns to every point x_i the measure $1/n$. The integral of any function f with respect to $\delta_{x,n}$ looks like

$$\delta_{x,n}[f] = \frac{f(x_1) + \dots + f(x_n)}{n}.$$

Denote by μ^n Cartesian power of a measure $\mu \in M(X)$, which is defined on X^n .

Theorem 3.1 (the local large deviations principle) *For any measure $\mu \in M(X)$, any functional $\nu \in B^*(X)$, and $\varepsilon > 0$ there exists a neighborhood $O(\nu)$ such that*

$$\mu^n \{x \in X^n \mid \delta_{x,n} \in O(\nu)\} \leq e^{-n(\rho(\nu, \mu) - \varepsilon)}. \quad (3.1)$$

On the other hand, for any $\varepsilon > 0$ and any neighborhood $O(\nu)$ the following asymptotic estimate holds:

$$\mu^n \{x \in X^n \mid \delta_{x,n} \in O(\nu)\} \geq e^{-n(\rho(\nu, \mu) + \varepsilon)}, \quad n \rightarrow \infty. \quad (3.2)$$

If $\rho(\nu, \mu) = +\infty$, then by the difference $\rho(\nu, \mu) - \varepsilon$ in (3.1) we mean an arbitrary positive number.

In the case of probability measure μ Theorem 3.1 is a partial case of Varadhan's large deviations principle (which explicit formulation can be found, e. g., in [4] and [14]). Therefore, this theorem can be deduced from Varadhan's large deviations principle by means of mere renormalization of μ . Nevertheless, we will prove it independently for the purpose of completeness.

Proof. By Theorem 2.2 for any $\varepsilon > 0$ there exists $\psi \in B(X)$ such that

$$\rho(\nu, \mu) - \varepsilon/2 < \nu[\psi] - \lambda(\psi, \mu). \quad (3.3)$$

Consider the probability measure $\mu_\psi = e^{\psi - \lambda(\psi, \mu)} \mu$. Obviously,

$$\frac{d\mu^n(x)}{d\mu_\psi^n(x)} = \prod_{i=1}^n \frac{d\mu(x_i)}{d\mu_\psi(x_i)} = \prod_{i=1}^n e^{\lambda(\psi, \mu) - \psi(x_i)} = e^{n(\lambda(\psi, \mu) - \delta_{x,n}[\psi])}. \quad (3.4)$$

Define a neighborhood of the functional ν as follows:

$$O(\nu) = \{\delta \in B^*(X) \mid \delta[\psi] > \nu[\psi] - \varepsilon/2\}.$$

Then it follows from (3.4) and (3.3) that under the condition $\delta_{x,n} \in O(\nu)$

$$\frac{d\mu^n(x)}{d\mu_\psi^n(x)} < e^{n(\lambda(\psi, \mu) - \nu[\psi] + \varepsilon/2)} < e^{n(-\rho(\nu, \mu) + \varepsilon)}.$$

Consequently,

$$\mu^n \{x \in X^n \mid \delta_{x,n} \in O(\nu)\} = \int_{\delta_{x,n} \in O(\nu)} d\mu^n(x) \leq \int_{\delta_{x,n} \in O(\nu)} e^{n(-\rho(\nu,\mu)+\varepsilon)} d\mu_\psi^n(x) \leq e^{-n(\rho(\nu,\mu)-\varepsilon)}.$$

Thus the first part of Theorem 3.1 is proved.

The estimate (3.2) is trivial if $\rho(\nu, \mu) = +\infty$. So it is enough to prove it only in the case when ν is a probability measure of the form $\nu = \varphi\mu$ and the Kullback action $\rho(\nu, \mu) = \nu[\ln \varphi]$ is finite. Fix any number $\varepsilon > 0$ and neighborhood $O(\nu)$. Define the sets

$$Y_n = \{x \in X^n \mid \delta_{x,n} \in O(\nu), \quad |\delta_{x,n}[\ln \varphi] - \nu[\ln \varphi]| < \varepsilon/2\}$$

(the last inequality in the braces means that $\varphi(x_i) > 0$ at each point of the sequence $x = (x_1, \dots, x_n)$). Note that for $x \in Y_n$

$$\frac{d\mu^n(x)}{d\nu^n(x)} = \prod_{i=1}^n \frac{d\mu(x_i)}{d\nu(x_i)} = \prod_{i=1}^n \frac{1}{\varphi(x_i)} = e^{-n\delta_{x,n}[\ln \varphi]} > e^{-n(\nu[\ln \varphi] + \varepsilon/2)}.$$

Consequently,

$$\mu^n(Y_n) = \int_{Y_n} d\mu^n(x) \geq \int_{Y_n} e^{-n(\nu[\ln \varphi] + \varepsilon/2)} d\nu^n(x) = e^{-n\rho(\nu,\mu) - n\varepsilon/2} \nu^n(Y_n). \quad (3.5)$$

By the Law of large numbers $\nu^n(Y_n) \rightarrow 1$. Hence (3.5) implies (3.2). \square

Corollary 3.2 (the McMillan theorem) *For any probability measure $\nu \in M_1(X)$ and $\varepsilon > 0$ there exists a neighborhood $O(\nu)$ such that*

$$\#\{x = (x_1, \dots, x_n) \in X^n \mid \delta_{x,n} \in O(\nu)\} \leq e^{n(H(\nu) + \varepsilon)}.$$

On the other hand, for any neighborhood $O(\nu)$ and $\varepsilon > 0$

$$\#\{x \in X^n \mid \delta_{x,n} \in O(\nu)\} \geq e^{n(H(\nu) - \varepsilon)} \quad \text{as } n \rightarrow \infty.$$

Here $H(\nu)$ denotes Shannon's entropy defined in (2.3).

Proof. This follows from equalities (2.2), (2.3), and the previous theorem, if we set $\mu(x) = 1$ for all $x \in X$. \square

4 Hausdorff dimension and the maximal dimension principle

Let us define the Hausdorff dimension of an arbitrary metric space Ω .

Suppose that Ω is covered by at most countable collection of subsets $\mathcal{U} = \{U_i\}$. Denote by $|\mathcal{U}|$ the diameter of this covering: $|\mathcal{U}| = \sup |U_i|$, where $|U_i|$ is the diameter of U_i . For every $\alpha \in \mathbb{R}$ put

$$\text{mes}(\mathcal{U}, \alpha) = \sum_i |U_i|^\alpha.$$

The *Hausdorff measure* (of dimension α) of the metric space Ω is

$$\text{mes}(\Omega, \alpha) = \lim_{|\mathcal{U}| \rightarrow 0} \text{mes}(\mathcal{U}, \alpha),$$

where \mathcal{U} is at most countable covering of Ω . Obviously,

$$\text{mes}(\mathcal{U}, \beta) \leq \text{mes}(\mathcal{U}, \alpha) |\mathcal{U}|^{\beta-\alpha} \quad \text{if } \beta \geq \alpha.$$

This implies the following property of the Hausdorff measure: if $\text{mes}(\Omega, \alpha) < \infty$ for some α , then $\text{mes}(\Omega, \beta) = 0$ for all $\beta > \alpha$.

The *Hausdorff dimension* of the space Ω is the number

$$\dim_H \Omega = \inf \{ \alpha \mid \text{mes}(\Omega, \alpha) = 0 \}. \quad (4.1)$$

In other words, $\dim_H \Omega = \alpha_0$ if $\text{mes}(\Omega, \alpha) = 0$ for all $\alpha > \alpha_0$ and $\text{mes}(\Omega, \alpha) = \infty$ for all $\alpha < \alpha_0$.

Below we will consider the space of sequences

$$X^{\mathbb{N}} = \{ x = (x_1, x_2, x_3, \dots) \}, \quad x_i \in X = \{1, \dots, r\}.$$

Let $x = (x_1, x_2, \dots) \in X^{\mathbb{N}}$. Denote by $Z_n(x)$ the set of sequences $y = (y_1, y_2, \dots)$ whose first n coordinates coincide with the same coordinates of x . This set will be called a *cylinder of rank n* . The collection of all cylinders generates the *Tychonoff topology* on the space $X^{\mathbb{N}}$ and the *cylinder σ -algebra* of subsets in $X^{\mathbb{N}}$.

Take an arbitrary positive function η on the set of all cylinders that possesses the following two properties: first, if $Z_n(x) \subset Z_m(y)$ then $\eta(Z_n(x)) \leq \eta(Z_m(y))$ and, second, $\eta(Z_n(x)) \rightarrow 0$ as $n \rightarrow \infty$ at each point $x \in X^{\mathbb{N}}$. Define the *cylinder metrics* on $X^{\mathbb{N}}$ by means of the formula

$$\text{dist}(x, y) = \eta(Z_n(x)), \quad \text{where } n = \max \{ m \mid Z_m(x) = Z_m(y) \}. \quad (4.2)$$

Evidently, the diameter of $Z_n(x)$ in this metrics coincides with $\eta(Z_n(x))$.

Suppose on $X^{\mathbb{N}}$, besides the cylinder metrics (4.2), a Borel measure μ is given. The function

$$d_\mu(x) = \lim_{n \rightarrow \infty} \frac{\ln \mu(Z_n(x))}{\ln |Z_n(x)|}$$

is called (*lower*) *pointwise dimension of the measure μ* .

The next theorem provides an effective tool for computing the Hausdorff dimensions of various subsets of $X^{\mathbb{N}}$.

Theorem 4.1 *Suppose $A \subset X^{\mathbb{N}}$. If there exists a finite Borel measure μ on $X^{\mathbb{N}}$ such that $d_\mu(x) \leq d$ for each point $x \in A$, then $\dim_H A \leq d$. On the contrary, if $d_\mu(x) \geq d$ for each $x \in A$ and the outer measure $\mu^*(A)$ is positive, then $\dim_H A \geq d$.*

It follows that if $d_\mu(x) \equiv d$ on the whole subset $A \subset X^{\mathbb{N}}$ then its dimension is equal to d .

A weakened version of the second part of Theorem 4.1 in which the condition $d_\mu(x) \geq d$ is replaced by the more strong one $\mu(Z_n(x)) \leq |Z_n(x)|^d$ is usually called the *mass distribution principle*.

Proof. Every cylinder $Z_n(x)$ is, in fact, a ball in the metrics (4.2), whose radius equals to its diameter, and vice versa, any ball in this metrics coincides with a cylinder. Besides, any two cylinders $Z_n(x)$ and $Z_m(y)$ either have empty intersection or one of them is embedded into other. Therefore, while computing the Hausdorff measure and dimension of a subset $A \subset X^{\mathbb{N}}$ it is enough to operate with only disjoint coverings of A by cylinders.

Suppose first that $d_\mu(x) < \alpha$ for all points $x \in A$. Then for each $x \in A$ there exist arbitrarily small cylinders $Z_n(x)$ satisfying the condition $|Z_n(x)|^\alpha < \mu(Z_n(x))$. Using this kind of cylinders we can put together a disjoint covering \mathcal{U} of the set A of arbitrarily small diameter. For this covering we have the inequalities

$$\text{mes}(\mathcal{U}, \alpha) = \sum_{Z_n(x) \in \mathcal{U}} |Z_n(x)|^\alpha \leq \sum_{Z_n(x) \in \mathcal{U}} \mu(Z_n(x)) \leq \mu(X^{\mathbb{N}}),$$

and hence $\dim_H A \leq \alpha$. Thus the first part of the theorem is proved.

Suppose now that $d_\mu(x) > \alpha$ for all points $x \in A$. Define the sets

$$A_\varepsilon = \{x \in A \mid |Z_n(x)|^\alpha > \mu(Z_n(x)) \text{ whenever } |Z_n(x)| < \varepsilon\}.$$

Obviously, $A = \bigcup_{\varepsilon > 0} A_\varepsilon$. Hence there exists an ε such that $\mu^*(A_\varepsilon) > 0$. Let \mathcal{U} be a disjoint covering of A by cylinders of diameters less than ε . From the definition of A_ε it follows that $\text{mes}(\mathcal{U}, \alpha) \geq \mu^*(A_\varepsilon)$. Therefore $\dim_H A \geq \alpha$, and thus the second part of the theorem is proved. \square

Theorem 4.1 was first proved by Billingsley in the case when the function η in (4.2) is a probability measure on $X^{\mathbb{N}}$ (see [3, Theorems 2.1 and 2.2]). An analog to this theorem for subsets $A \subset \mathbb{R}^r$ was proved in [15] and [13].

Each point $x = (x_1, x_2, \dots) \in X^{\mathbb{N}}$ generates a sequence of empirical measures $\delta_{x,n}$ on the set X :

$$\delta_{x,n}(i) = \frac{\#\{t \mid x_t = i, t \leq n\}}{n}, \quad i \in X.$$

In other words, $\delta_{x,n}(i)$ is the fraction of those coordinates of the vector (x_1, \dots, x_n) that coincide with i .

For every probability measure $\nu \in M_1(X)$ let us define its *basin* $B(\nu)$ as the set of all points $x \in X^{\mathbb{N}}$ such that $\delta_{x,n}$ converges to ν .

Evidently, basins of different measures do not intersect each other and are nonempty. If $x \in B(\nu)$, and $y \in X^{\mathbb{N}}$ differs from x in only finite number of coordinates, then $y \in B(\nu)$. This implies density of each basin in $X^{\mathbb{N}}$.

Every measure $\nu \in M_1(X)$ generates Bernoulli distribution $P_\nu = \nu^{\mathbb{N}}$ on the space $X^{\mathbb{N}}$. By the strong law of large numbers the basin $B(\nu)$ has probability one with respect to Bernoulli distribution P_ν , and its complement has zero probability P_ν . In particular, any basin different from $B(\nu)$ has zero probability.

Points that does not belong to the union of all basins will be called *irregular*. The set of irregular points has zero probability with respect to any distribution P_ν , where $\nu \in M_1(X)$. As a result, $X^{\mathbb{N}}$ turns out to be decomposed into the disjoint union of different basins and the set of irregular points.

Let us fix some numbers $\theta(i) \in (0, 1)$ for all elements $i \in X = \{1, \dots, r\}$, and define a *cylinder θ -metrics* on $X^{\mathbb{N}}$ by the rule

$$\text{dist}(x, y) = \prod_{t=1}^n \theta(x_t), \quad n = \inf \{t \mid x_t \neq y_t\} - 1. \quad (4.3)$$

It is a partial case of the cylinder metrics (4.2).

For each measure $\nu \in M_1(X)$ and θ -metrics (4.3) define the quantity

$$S(\nu, \theta) = \frac{\sum_{i=1}^r \nu(i) \ln \nu(i)}{\sum_{i=1}^r \nu(i) \ln \theta(i)}. \quad (4.4)$$

We will call it the *Billingsley entropy* because he was the first who wrote down this formula and applied it for the computation of Hausdorff dimensions [2]. He expressed also this quantity in terms of Shannon's entropy and the Kullback action:

$$S(\nu, \theta) = \frac{H(\nu)}{H(\nu) + \rho(\nu, \theta)}.$$

Theorem 4.2 *Hausdorff dimension of any basin $B(\nu)$ relative to the θ -metrics (4.3) is equal to the Billingsley entropy $S(\nu, \theta)$.*

A partial case of this theorem in which $\theta(1) = \dots = \theta(r) = 1/r$ was first proved by Eggleston [5]. In the complete form this theorem and its generalizations were proved by Billingsley in [2, 3].

Proof. Assume first that $\nu(i) > 0$ for every $i = 1, \dots, r$. Obviously,

$$\frac{\ln P_\nu(Z_n(x))}{\ln |Z_n(x)|} = \frac{\sum_{t=1}^n \ln \nu(x_t)}{\sum_{t=1}^n \ln \theta(x_t)} = \frac{\sum_{i=1}^r n \delta_{x,n}(i) \ln \nu(i)}{\sum_{i=1}^r n \delta_{x,n}(i) \ln \theta(i)}.$$

Hence for each point $x \in B(\nu)$ we have

$$d_{P_\nu}(x) = \lim_{n \rightarrow \infty} \frac{\ln P_\nu(Z_n(x))}{\ln |Z_n(x)|} = \frac{\sum_{i=1}^r \nu(i) \ln \nu(i)}{\sum_{i=1}^r \nu(i) \ln \theta(i)} = S(\nu, \theta). \quad (4.5)$$

Applying Theorem 4.1 to the set $A = B(\nu)$ and measure $\mu = P_\nu$, we obtain the statement of Theorem 4.2.

In the general case the same argument provides only lower bound $d_{P_\nu}(x) \geq S(\nu, \theta)$, that implies the lower bound $\dim_H B(\nu) \geq S(\nu, \theta)$. The inverse inequality is provided by the next lemma. \square

Lemma 4.3 *Suppose the space $X^\mathbb{N}$ is equipped with the metrics (4.3). Then for any measure $\nu \in M_1(X)$ and $\varepsilon > 0$ there exists a neighborhood $O(\nu)$ such that Hausdorff dimension of the set*

$$A = \{x \in X^\mathbb{N} \mid \forall N \exists n > N: \delta_{x,n} \in O(\nu)\}$$

does not exceed $S(\nu, \theta) + \varepsilon$.

Proof. Fix a measure $\nu \in M_1(X)$ and an arbitrary positive number \varkappa . By McMillan's theorem there exists a neighborhood $O(\nu)$ such that for each positive integer n

$$\#\{Z_n(x) \mid \delta_{x,n} \in O(\nu)\} \leq e^{n(H(\nu) + \varkappa)}. \quad (4.6)$$

Decrease this neighborhood in such a way that, in addition, for every measure $\delta \in O(\nu)$ the next inequality holds:

$$\sum_{i=1}^r \delta(i) \ln \theta(i) < \sum_{i=1}^r \nu(i) \ln \theta(i) + \varkappa.$$

Then for every cylinder $Z_n(x)$ satisfying the condition $\delta_{x,n} \in O(\nu)$ we have the estimate

$$\begin{aligned} |Z_n(x)| &= \prod_{t=1}^n \theta(x_t) = \exp \left\{ \sum_{t=1}^n \ln \theta(x_t) \right\} = \exp \left\{ n \sum_{i=1}^r \delta_{x,n}(i) \ln \theta(i) \right\} < \\ &< \exp \left\{ n \sum_{i=1}^r \nu(i) \ln \theta(i) + n\kappa \right\}. \end{aligned} \quad (4.7)$$

For any positive integer N the set A is covered by the collection of cylinders

$$\mathcal{U}_N = \bigcup_{n=N}^{\infty} \{Z_n(x) \mid \delta_{x,n} \in O(\nu)\}.$$

Evidently, the diameter of this covering goes to zero when N increases. Now we can evaluate $\text{mes}(\mathcal{U}_N, \alpha)$ by means of formulas (4.6) and (4.7):

$$\begin{aligned} \text{mes}(\mathcal{U}_N, \alpha) &= \sum_{Z_n(x) \in \mathcal{U}_N} |Z_n(x)|^\alpha \leq \sum_{n=N}^{\infty} e^{n(H(\nu) + \kappa)} \exp \left\{ \alpha n \sum_{i=1}^r \nu(i) \ln \theta(i) + \alpha n \kappa \right\} = \\ &= \sum_{n=N}^{\infty} \exp \left\{ n \left(- \sum_{i=1}^r \nu(i) \ln \nu(i) + \alpha \sum_{i=1}^r \nu(i) \ln \theta(i) + \kappa + \alpha \kappa \right) \right\}. \end{aligned} \quad (4.8)$$

If $\alpha > S(\nu, \theta)$, then we can choose so small $\kappa > 0$ that the last exponent in braces is negative, and all the sum (4.8) goes to zero as $N \rightarrow \infty$. Therefore Hausdorff measure (of dimension α) of the set A is zero, and hence $\dim_H A$ does not exceed α . \square

We will say that a sequence of empirical measures $\delta_{x,n}$ *condenses* on a subset $V \subset M_1(X)$ (notation $\delta_{x,n} \succ V$) if it has at least one limit point in V .

Similarly to the famous large deviations principle by Varadhan [4, 14], it is natural that the next theorem be named the *maximal dimension principle*.

Theorem 4.4 *Let the space $X^{\mathbb{N}}$ be equipped with the cylinder θ -metrics (4.3). Then for any nonempty subset $V \subset M_1(X)$*

$$\dim_H \{x \in X^{\mathbb{N}} \mid \delta_{x,n} \succ V\} = \sup_{\nu \in V} S(\nu, \theta). \quad (4.9)$$

Proof. The set $A = \{x \in X^{\mathbb{N}} \mid \delta_{x,n} \succ V\}$ contains basins of all measures $\nu \in V$. So by Theorem 4.2 its dimension is not less than the right hand side of (4.9).

It is easily seen from the definition (4.4) of the Billingsly entropy $S(\nu, \theta)$ that it depends continuously on the measure $\nu \in M_1(X)$. Consider the closure \bar{V} of V . Obviously, it is compact. Fix any $\varepsilon > 0$. By Lemma 4.3 for any measure $\nu \in \bar{V}$ there exists a neighborhood $O(\nu)$ such that

$$\dim_H \{x \in X^{\mathbb{N}} \mid \delta_{x,n} \succ O(\nu)\} \leq S(\nu, \theta) + \varepsilon \leq \sup_{\nu \in V} S(\nu, \theta) + \varepsilon. \quad (4.10)$$

Pick out a finite covering of \bar{V} composed of neighborhoods of this sort. Then the set $A = \{x \in X^{\mathbb{N}} \mid \delta_{x,n} \succ V\}$ will be covered by a finite collection of sets of the form $\{x \in X^{\mathbb{N}} \mid \delta_{x,n} \succ O(\nu)\}$ satisfying (4.10). By the arbitrariness of ε this implies the statement of Theorem 4.4. \square

A very similar to Theorem 4.4 result was proved by Billingsley in [2, Theorem 7.1].

Suppose that a certain subset $\Xi \subset X$ is specified in the set $X = \{1, \dots, r\}$. In this case the subset $\Xi^{\mathbb{N}} \subset X^{\mathbb{N}}$ will be named the *generalized Cantor set*. It consists of those sequences $x = (x_1, x_2, \dots)$ in which all $x_t \in \Xi$.

Theorem 4.5 *If the space $X^{\mathbb{N}}$ is equipped with the θ -metrics (4.3) then Hausdorff dimension of the generalized Cantor set $\Xi^{\mathbb{N}}$ coincides with the unique solution of Moran's equation*

$$\sum_{i \in \Xi} \theta(i)^s = 1. \quad (4.11)$$

This theorem was first proved by Moran in 1946 [12] for generalized Cantor subsets of the real axis and afterwards it was extended by Hutchinson [10] to the attractors of self-similar geometric constructions in \mathbb{R}^r . Let us show how it can be derived from the maximal dimension principle.

Proof. Let s be the solution to Moran's equation. Introduce a probability distribution ν on X , setting $\nu(i) = \theta(i)^s$ for $i \in \Xi$ and $\nu(i) = 0$ for $i \notin \Xi$. Then

$$S(\nu, \theta) = \frac{\sum_{i=1}^r \nu(i) \ln \nu(i)}{\sum_{i=1}^r \nu(i) \ln \theta(i)} = \frac{\sum_{i \in \Xi} \theta(i)^s \ln \theta(i)^s}{\sum_{i \in \Xi} \theta(i)^s \ln \theta(i)} = s. \quad (4.12)$$

Consider the set $B(\nu) \cap \Xi^{\mathbb{N}}$. It has the unit measure with respect to the distribution $P_\nu = \nu^{\mathbb{N}}$. Besides, for every point $x \in B(\nu) \cap \Xi^{\mathbb{N}}$ by (4.5) we have the equality $d_{P_\nu}(x) = S(\nu, \theta)$. In this setting it follows from Theorem 4.1 and formula (4.12) that

$$\dim_H(B(\nu) \cap \Xi^{\mathbb{N}}) = S(\nu, \theta) = s. \quad (4.13)$$

Denote by V the collection of all probability measures on X supported on $\Xi \subset X$. Evidently, for each point $x \in \Xi^{\mathbb{N}}$ all the limit points of the sequence $\delta_{x,n}$ belong to V . Hence, we can apply Theorem 4.4 that implies

$$\dim_H \Xi^{\mathbb{N}} \leq \dim_H \{x \in X^{\mathbb{N}} \mid \delta_{x,n} \succ V\} = \sup_{\nu \in V} S(\nu, \theta) = \sup_{\nu \in V} \frac{\sum_{i \in \Xi} \nu(i) \ln \nu(i)}{\sum_{i \in \Xi} \nu(i) \ln \theta(i)}. \quad (4.14)$$

Note that for every measure $\nu \in V$

$$s \sum_{i \in \Xi} \nu(i) \ln \theta(i) - \sum_{i \in \Xi} \nu(i) \ln \nu(i) = \sum_{\nu(i) > 0} \nu(i) \ln \frac{\theta(i)^s}{\nu(i)} \leq \ln \left\{ \sum_{\nu(i) > 0} \nu(i) \frac{\theta(i)^s}{\nu(i)} \right\} \leq 0,$$

where we have used concavity of the logarithm function. It follows that the right hand side in (4.14) does not exceed s . Finally, comparing (4.13) and (4.14), we obtain the desired equality $\dim_H \Xi^{\mathbb{N}} = s$. \square

5 Branching processes

First let us introduce the basic notions about the simplest Galton–Watson branching process.

Suppose that a random variable Z takes nonnegative values $k \in \mathbb{Z}_+$ with probabilities p_k . The *Galton–Watson branching process* is a sequence of integer-valued random

variables Z_0, Z_1, Z_2, \dots such that $Z_0 \equiv 1$, $Z_1 = Z$, and further each Z_{n+1} is defined as the sum of Z_n independent counterparts of the random variable Z . In particular, if $Z_n = 0$ then $Z_{n+1} = 0$ as well. Usually Z_n is thought of as the total number of descendants in n -th generation of a unique common ancestor under the condition that each descendant independently of others gives birth to Z children.

It is known that in some cases the posterity of the initial ancestor may degenerate (when starting from a certain n all Z_n are zeros) and in other cases it can “flourish” (when Z_n grows exponentially). The type of behavior of the brunching process depends on the mean number of children of any individual

$$m = \mathbb{E}Z = \sum_{k=0}^{\infty} kp_k$$

and on the generating function of that number

$$f(s) = f_1(s) = \sum_{k=0}^{\infty} p_k s^k.$$

Obviously, the restriction of the function $f(s)$ to the segment $[0, 1]$ is nonnegative, nondecreasing, convex, and satisfies $f(1) = 1$ and $f'(1) = m$.

In the theory of brunching processes (see, for instance, [1, 9]) the following statements have been proved.

Theorem 5.1 *The generating functions of the number of descendants in n -th generation*

$$f_n(s) = \sum_{k=0}^{\infty} \mathbb{P}\{Z_n = k\} s^k$$

satisfy the recursion relation $f_{n+1}(s) = f(f_n(s))$.

Theorem 5.2 *If $m \leq 1$ then the brunching process degenerates almost surely (except the case when each individual gives birth to exactly one child). If $m > 1$ then the probability q of degeneration is less than 1 and coincides with a unique nonunit root of the equation $f(s) = s$ on the segment $[0, 1]$.*

Theorem 5.3 *If $m > 1$ and $\mathbb{E}Z^2 < \infty$ then the sequence $W_n = Z_n/m^n$ converges almost surely to a random variable W such that $\mathbb{P}\{W > 0\} = 1 - q$. If $m > 1$ and $\mathbb{E}Z^2 = \infty$ then for any number $m' < m$ with probability $1 - q$*

$$\lim_{n \rightarrow \infty} Z_n/m^n < \infty, \quad \lim_{n \rightarrow \infty} Z_n/(m')^n = \infty$$

(here q is the probability of degeneration of the brunching process).

Thereby, in the case $m > 1$ there is an alternative for the total number of descendants Z_n : either it vanishes at a certain moment n_0 (with probability $q < 1$) or it is asymptotically equivalent to Wm^n (with the complementary probability $1 - q$), where the random variable $W > 0$ does not depend on n (except the case $\mathbb{E}Z^2 = \infty$, when only the logarithmic equivalence $\ln Z_n \sim \ln m^n$ is guaranteed). All other types of the descendants' number behavior have zero probability.

We will exploit these theorems in the study of colored brunching processes.

Suppose now that each individual may give birth to children of r different colors (or r different genders, if one likes). We will suppose that the posterity of each individual in the first generation represents a random set X containing random number k_1 of children of the first color, random number k_2 of children of the second color, and so on up to k_r children of color r . All elements of X (including elements of the same color) are treated as different. The ordered array $k = (k_1, k_2, \dots, k_r) \in \mathbb{Z}_+^r$ will be called the *color structure* of the set of children X . Denote by p_k the probability of birth of the set X with color structure $k = (k_1, k_2, \dots, k_r)$. Naturally, all the probabilities p_k are nonnegative and

$$\sum_{k \in \mathbb{Z}_+^r} p_k = 1.$$

If an individual x_1 gave birth to x_2 , then x_2 gave birth to x_3 , and so on up to an individual x_n , then the sequence $x = (x_1, \dots, x_n)$ will be called the *genetic line* of length n .

Let us construct a new branching process taking into account not only the total number of descendants but also the color of each individual and all its upward and downward lineal relations. This process may be thought of as a random *genealogical tree* with a common ancestor in the root and all its descendants in the vertices, where each parent is linked with all its children. In the case of degenerating population its genealogical tree is finite, and in the case of “flourishing” one the tree is infinite.

Formally it is convenient to define such a process as a sequence of random sets X_n containing all genetic lines of length n . As the first set X_1 we take X . The subsequent X_n are built up by induction: if X_n is already known, then for all genetic lines $(x_1, \dots, x_n) \in X_n$ define disjoint independent random sets of children $X(x_1, \dots, x_n)$, each with color structure distribution as in X , and put

$$X_{n+1} = \{(x_1, \dots, x_n, x_{n+1}) \mid (x_1, \dots, x_n) \in X_n, x_{n+1} \in X(x_1, \dots, x_n)\}.$$

The built in such a way stochastic process X_1, X_2, X_3, \dots will be referred to as the *colored branching process* (or *unconditional colored branching process* if one wishes to emphasize that the posterity of any individual is independent of its color and genealogy).

6 The McMillan theorem for colored branching processes

Consider a colored branching process X_1, X_2, \dots determined by a finite collection of colors $\Omega = \{1, \dots, r\}$ and a probability distribution $\{p_k \mid k \in \mathbb{Z}_+^r\}$, where $k = (k_1, \dots, k_r)$ is the color structure of each individual's set of children X . We will always think that X_1 is generated by a unique initial individual.

For any genetic line $x = (x_1, \dots, x_n) \in X_n$ define the *spectrum* $\delta_{x,n}$ as the corresponding empirical measure on Ω by the rule

$$\delta_{x,n}(i) = \frac{\#\{t \mid g(x_t) = i\}}{n}, \quad i \in \Omega, \quad (6.1)$$

where $g(x_t)$ denotes the color of x_t . In other words, $\delta_{x,n}(i)$ is the fraction of individuals of color i in the genetic line x . Our next goal is to obtain asymptotical estimates for cardinalities of the random sets

$$\{x = (x_1, \dots, x_n) \in X_n \mid \delta_{x,n} \in O(\nu)\},$$

where $O(\nu)$ is a small neighborhood of the distribution ν on the set of colors Ω .

Denote by $\mu(i)$ the expectation of members of color i in X :

$$\mu(i) = \sum_{k \in \mathbb{Z}_+^r} k_i p_k, \quad i = 1, \dots, r. \quad (6.2)$$

Provided all $\mu(i)$ are finite, the vector $\mu = (\mu(1), \dots, \mu(r))$ can be regarded as a measure on the set of colors Ω . This measure generates the measure μ^n on Ω^n as Cartesian product.

Define a mapping $G : X_n \rightarrow \Omega^n$ by means of the formula

$$G(x_1, \dots, x_n) = (g(x_1), \dots, g(x_n)),$$

where $g(x_t)$ is the color of x_t .

Lemma 6.1 *For any $\omega = (\omega_1, \dots, \omega_n) \in \Omega^n$ we have*

$$\mathbb{E} \# \{x \in X_n \mid G(x) = \omega\} = \prod_{t=1}^n \mu(\omega_t) = \mu^n(\omega). \quad (6.3)$$

Proof. Cast out the last coordinate in ω and let $\omega' = (\omega_1, \dots, \omega_{n-1})$. For any genetic line $(x_1, \dots, x_{n-1}) \in X_{n-1}$, by virtue of the definition of unconditional colored branching process we have

$$\mathbb{E} \# \{x_n \in X(x_1, \dots, x_{n-1}) \mid g(x_n) = \omega_n\} = \mu(\omega_n).$$

Evidently, this expression does not depend on $x' = (x_1, \dots, x_{n-1})$. Therefore,

$$\mathbb{E} \# \{x \in X_n \mid G(x) = \omega\} = \mathbb{E} \# \{x' \in X_{n-1} \mid G(x') = \omega'\} \mu(\omega_n).$$

Repeated application of the latter equality gives (6.3). \square

Define for the measure μ from (6.2) the Kullback action

$$\rho(\nu, \mu) = \sum_{i \in \Omega} \nu(i) \ln \frac{\nu(i)}{\mu(i)}, \quad \nu \in M_1(\Omega),$$

where $M_1(\Omega)$ is the set of all probability measures on Ω . This formula is a copy of (2.2).

Theorem 6.2 *Suppose X_1, X_2, \dots is an unconditional colored brunching process with finite collection of colors Ω . Then for any $\varepsilon > 0$ and probability measure $\nu \in M_1(\Omega)$ there exists a neighborhood $O(\nu) \subset M_1(\Omega)$ such that for all natural n*

$$\mathbb{E} \# \{x \in X_n \mid \delta_{x,n} \in O(\nu)\} \leq e^{n(-\rho(\nu, \mu) + \varepsilon)}. \quad (6.4)$$

On the other hand, for any $\varepsilon > 0$ and any neighborhood $O(\nu)$

$$\mathbb{E} \# \{x \in X_n \mid \delta_{x,n} \in O(\nu)\} \geq e^{n(-\rho(\nu, \mu) - \varepsilon)} \quad \text{as } n \rightarrow \infty. \quad (6.5)$$

If $\rho(\nu, \mu) = +\infty$, the expression $-\rho(\nu, \mu) + \varepsilon$ in (6.4) should be treated as an arbitrary negative real number.

Proof. It follows from (6.1) that for every genetic line $x \in X_n$ its spectrum $\delta_{x,n}$ coincides with the empirical measure $\delta_{\omega,n}$, where $\omega = G(x)$. Therefore,

$$\#\{x \in X_n \mid \delta_{x,n} \in O(\nu)\} = \sum_{\omega \in \Omega^n: \delta_{\omega,n} \in O(\nu)} \#\{x \in X_n \mid G(x) = \omega\}. \quad (6.6)$$

It follows from (6.3) and (6.6) that

$$\mathbf{E} \#\{x \in X_n \mid \delta_{x,n} \in O(\nu)\} = \mu^n \{\omega \in \Omega^n \mid \delta_{\omega,n} \in O(\nu)\}.$$

The latter equality converts estimates (6.4) and (6.5) into already proved estimates (3.1), (3.2) from the large deviations principle. \square

Remarkable that the last reference to the large deviations principle serves a unique “umbilical cord” linking the first three sections of the paper with others.

Now we are ready to state an analog of the McMillan theorem for colored branching processes. Let q^* be a probability of degeneration of the process (probability of the occasion that starting from a certain number n all the sets X_n turn out to be empty).

Theorem 6.3 *Suppose X_1, X_2, \dots is an unconditional colored brunching process with finite collection of colors Ω . Then for any $\varepsilon > 0$ and any probability measure $\nu \in M_1(\Omega)$ there exists a neighborhood $O(\nu) \subset M_1(\Omega)$ such that for almost sure*

$$\#\{x \in X_n \mid \delta_{x,n} \in O(\nu)\} < e^{n(-\rho(\nu, \mu) + \varepsilon)} \quad \text{as } n \rightarrow \infty. \quad (6.7)$$

On the other hand, if $\rho(\nu, \mu) < 0$ then for any neighborhood $O(\nu)$ and positive ε the estimate

$$\#\{x \in X_n \mid \delta_{x,n} \in O(\nu)\} > e^{n(-\rho(\nu, \mu) - \varepsilon)} \quad \text{as } n \rightarrow \infty \quad (6.8)$$

holds with probability $1 - q^*$ (or almost surely under the condition that our branching process does not degenerate).

Proof. Application of Chebyshev’s inequality to (6.4) gives

$$\mathbf{P}\{\#\{x \in X_n \mid \delta_{x,n} \in O(\nu)\} \geq e^{n(-\rho(\nu, \mu) + 2\varepsilon)}\} \leq e^{-n\varepsilon}.$$

Sum up these inequalities over all $n \geq N$:

$$\mathbf{P}\{\exists n \geq N: \#\{x \in X_n \mid \delta_{x,n} \in O(\nu)\} \geq e^{n(-\rho(\nu, \mu) + 2\varepsilon)}\} \leq \frac{e^{-N\varepsilon}}{1 - e^{-\varepsilon}}.$$

This implies (6.7) with constant 2ε instead of ε , that does not change its sense.

Proceed to the second part of the theorem. Let $\varkappa = -\rho(\nu, \mu) - \varepsilon$ and the number ε be so small that $\varkappa > 0$. By the second part of Theorem 6.2 for any neighborhood $O(\nu)$ there exists N such that

$$\mathbf{E} \#\{x \in X_N \mid \delta_{x,N} \in O(\nu)\} > e^{N\varkappa}.$$

Without loss of generality we may assume that $O(\nu)$ is convex.

Construct a Galton–Watson branching process satisfying the conditions

$$Z_1 = \#\{x \in X_N \mid \delta_{x,N} \in O(\nu)\}, \quad (6.9)$$

$$Z_n \leq \#\{x \in X_{nN} \mid \delta_{x,nN} \in O(\nu)\}, \quad n = 2, 3, \dots \quad (6.10)$$

Let the random variable Z_1 be defined by (6.9). For $n > 1$ define Z_n as a total number of genetic lines

$$(x_1, \dots, x_N, \dots, x_{(n-1)N+1}, \dots, x_{nN}) \in X_{nN}$$

such that the spectrum of each segment $(x_{kN+1}, \dots, x_{(k+1)N})$ belongs to $O(\nu)$. In other words, we will treat as “individuals” of the process Z_1, Z_2, Z_3, \dots those segments $(x_{kN+1}, \dots, x_{(k+1)N})$ of genetic lines of the initial process whose spectrum lies in $O(\nu)$. Then (6.10) follows from convexity of $O(\nu)$, and from unconditionality of the initial colored branching process it can be concluded that the sequence Z_1, Z_2, \dots in fact forms a Galton–Watson branching process.

By construction, $\mathbb{E}Z_1 > e^{N\kappa}$. In this setting Theorem 5.3 asserts that there is an alternative for the sequence Z_n : either it tends to zero with a certain probability $q < 1$ or it grows faster than $e^{nN\kappa}$ with probability $1 - q$. In the second case, by virtue of (6.10),

$$\#\{x \in X_{nN} \mid \delta_{x,nN} \in O(\nu)\} e^{-nN\kappa} \rightarrow \infty \quad n \rightarrow \infty. \quad (6.11)$$

To finish the proof we have to do two things: verify that in fact (6.11) is valid with probability $1 - q^*$ and get rid of the multiplier N there. To do this we will exploit two ideas. First, if the colored branching process X_1, X_2, \dots were generated by m initial individuals instead of the unique one, then (6.11) would be valid with probability at least $1 - q^m$. Second, if one genetic line is a part of another and the ratio of their lengths is close to 1 then their spectra are close as well.

Obviously, the total number of individuals in the n -th generation of the initial branching process X_1, X_2, X_3, \dots equals $|X_n|$. The sequence of random variables $|X_n|$ forms a Galton–Watson branching process with probability of degeneration q^* , that does not exceed q . Therefore, the sequence $|X_n|$ grows exponentially with probability $1 - q^*$.

Consider the colored brunching process $X_{k+1}, X_{k+2}, X_{k+3}, \dots$ obtained from the initial one by virtue of truncation of the first k generations. It represents a union of $|X_k|$ independent brunching processes generated by all individuals of k -th generation. It satisfies (6.11) with probability at least $1 - q^{|X_k|}$. Hence for the initial process with even greater probability we obtain the condition

$$\#\{x \in X_{k+nN} \mid \delta_{x,k+nN} \in O^*(\nu)\} e^{-nN\kappa} \rightarrow \infty \quad n \rightarrow \infty, \quad (6.12)$$

where $O^*(\nu)$ is an arbitrary neighborhood of ν containing the closure of $O(\nu)$.

Suppose the sequence $|X_n|$ grows exponentially. Then for every $m \in \mathbb{N}$ define the numbers

$$k_i = \min\{k : |X_k| \geq m, \quad k = i \bmod N\}, \quad i = 0, 1, \dots, N-1.$$

For each $k = k_i$, the condition (6.12) holds with probability at least $1 - q^m$, and in common they give the estimate

$$\#\{x \in X_n \mid \delta_{x,n} \in O^*(\nu)\} > e^{n\kappa} \quad \text{as } n \rightarrow \infty$$

with probability at least $1 - Nq^m$. By virtue of the arbitrariness of m this estimate is valid almost surely (under the condition $|X_n| \rightarrow \infty$, which takes place with probability $1 - q^*$). It is equivalent to (6.8). \square

7 Dimensions of random fractals (upper bounds)

We proceed investigation of the colored brunching process X_1, X_2, \dots with finite collection of colors $\Omega = \{1, \dots, r\}$. Let us consider the corresponding set of infinite genetic lines

$$X_\infty = \{x = (x_1, x_2, x_3, \dots) \mid (x_1, \dots, x_n) \in X_n \ \forall n \in \mathbb{N}\}.$$

Define the cylinder θ -metrics on X_∞

$$\text{dist}(x, y) = \prod_{t=1}^n \theta(x_t), \quad n = \inf \{t \mid x_t \neq y_t\} - 1, \quad (7.1)$$

where the numbers $\theta(1), \dots, \theta(r)$ are taken from $(0, 1)$.

We will be interested in Hausdorff dimensions of both the space X_∞ and its various subsets defined in terms of partial limits of empirical measures on Ω (those measures are called spectra and denoted $\delta_{x,n}$). If the colored brunching process degenerates then X_∞ is empty. Therefore of interest is only the case when $m = \mathbf{E}|X_1| > 1$ and the cardinality of X_n increases with rate of order m^n .

As before, denote by $\mu(i)$, where $i \in \Omega$, the expectation of individuals of color i in the random set X_1 . It will be always supposed that $\mu(i) < \infty$. Consider any probability measure $\nu \in M_1(\Omega)$. It will be proved below that the dimension of $\{x \in X_\infty \mid \delta_{x,n} \rightarrow \nu\}$ can be computed by means of the function

$$d(\nu, \mu, \theta) = \frac{\rho(\nu, \mu)}{\sum_{i=1}^r \nu(i) \ln \theta(i)} = \frac{\sum_{i=1}^r \nu(i) \ln \frac{\nu(i)}{\mu(i)}}{\sum_{i=1}^r \nu(i) \ln \theta(i)}. \quad (7.2)$$

We will name it the *Billingsley–Kullback entropy*.

In (7.2) the numerator is the Kullback action and the denominator is negative. If μ is a probability measure on Ω then the Kullback action is nonnegative. But in our setting this is not the case since $m = \mu(1) + \dots + \mu(r) > 1$. In particular, if $\mu(i) > \nu(i)$ for all $i \in \Omega$ then the Kullback action will be negative, and the Billingsley–Kullback entropy positive. Note, in addition, that if $\mu(1) = \dots = \mu(r) = 1$ then $-\rho(\nu, \mu)$ is equal to Shannon's entropy $H(\nu)$, and the whole of Billingsley–Kullback entropy turns into the Billingsley entropy (4.4).

Lemma 7.1 *Let the space X_∞ of infinite genetic lines be equipped with the metrics (7.1). Then for any probability measure $\nu \in M_1(\Omega)$ and any $\varepsilon > 0$ there exists a neighborhood $O(\nu)$ such that Hausdorff dimension of the set*

$$A = \{x \in X_\infty \mid \forall N \exists n > N: \delta_{x,n} \in O(\nu)\}$$

does not exceed $d(\nu, \mu, \theta) + \varepsilon$ almost surely.

Proof is carried out in the same manner as in Lemma 4.3. Take any $\varkappa > 0$. By Theorem 6.4 there exists a neighborhood $O(\nu)$ such that almost surely

$$\#\{x \in X_n \mid \delta_{x,n} \in O(\nu)\} \leq e^{n(-\rho(\nu,\mu)+\varkappa)} \quad \text{as } n \rightarrow \infty. \quad (7.3)$$

Reduce this neighborhood in such a way that in addition for all measures $\delta \in O(\nu)$,

$$\sum_{i=1}^r \delta(i) \ln \theta(i) < \sum_{i=1}^r \nu(i) \ln \theta(i) + \varkappa.$$

Then for each cylinder $Z_n(x)$ satisfying the condition $\delta_{x,n} \in O(\nu)$ we have the estimate

$$\begin{aligned} |Z_n(x)| &= \prod_{t=1}^n \theta(x_t) = \exp \left\{ \sum_{t=1}^n \ln \theta(x_t) \right\} = \exp \left\{ n \sum_{i=1}^r \delta_{x,n}(i) \ln \theta(i) \right\} < \\ &< \exp \left\{ n \sum_{i=1}^r \nu(i) \ln \theta(i) + n\varkappa \right\}. \end{aligned} \quad (7.4)$$

For every natural N the set A is covered by the collection of cylinders

$$\mathcal{U}_N = \bigcup_{n=N}^{\infty} \{Z_n(x) \mid \delta_{x,n} \in O(\nu)\}.$$

Evidently, the diameter of this covering tends to zero as $N \rightarrow \infty$. Hence $\text{mes}(\mathcal{U}_N, \alpha)$ can be estimated by virtue of formulas (7.3) and (7.4):

$$\begin{aligned} \text{mes}(\mathcal{U}_N, \alpha) &= \sum_{Z_n(x) \in \mathcal{U}_N} |Z_n(x)|^\alpha \leq \sum_{n=N}^{\infty} e^{n(-\rho(\nu,\mu)+\varkappa)} \exp \left\{ \alpha n \sum_{i=1}^r \nu(i) \ln \theta(i) + \alpha n \varkappa \right\} = \\ &= \sum_{n=N}^{\infty} \exp \left\{ n \left(- \sum_{i=1}^r \nu(i) \ln \frac{\nu(i)}{\mu(i)} + \alpha \sum_{i=1}^r \nu(i) \ln \theta(i) + \varkappa + \alpha \varkappa \right) \right\}. \end{aligned} \quad (7.5)$$

If $\alpha > d(\nu, \mu, \theta)$ then \varkappa can be chosen so small that the last exponent in braces is negative, and all the sum (7.5) tends to zero as $N \rightarrow \infty$. Therefore Hausdorff measure (of dimension α) of the set A is zero, and its dimension does not exceed α . \square

As before, we say that the sequence of empirical measures $\delta_{x,n}$ condenses on a subset $V \subset M_1(\Omega)$ (notation $\delta_{x,n} \succ V$) if it has a limit point in V .

Theorem 7.2 *Let X_1, X_2, X_3, \dots be an unconditional colored brunching process with finite set of colors Ω , and the set X_∞ of all infinite genetic lines equipped with the cylinder metrics (7.1). Then for any subset $V \subset M_1(\Omega)$ almost surely*

$$\dim_H \{x \in X_\infty \mid \delta_{x,n} \succ V\} \leq \sup_{\nu \in V} d(\nu, \mu, \theta). \quad (7.6)$$

In particular, $\dim_H X_\infty \leq s$ for almost sure, where s is a unique root of the ‘‘Bowen equation’’

$$\sum_{i=1}^r \mu(i) \theta(i)^s = 1. \quad (7.7)$$

Proof. It follows from the definition of the Billingsley–Kullback entropy $d(\nu, \mu, \theta)$ that it depends continuously on the measure $\nu \in M_1(\Omega)$. Let \bar{V} be the closure of V . Obviously, it is compact. Take an arbitrary $\varepsilon > 0$. By Lemma 7.1 for any measure $\nu \in \bar{V}$ there exists a neighborhood $O(\nu)$ such that almost surely

$$\dim_H \{x \in X_\infty \mid \delta_{x,n} \succ O(\nu)\} \leq d(\nu, \mu, \theta) + \varepsilon \leq \sup_{\nu \in V} d(\nu, \mu, \theta) + \varepsilon. \quad (7.8)$$

Choose a finite covering of \bar{V} by neighborhoods of this kind. Then the set $\{x \in X_\infty \mid \delta_{x,n} \succ V\}$ will be covered by a finite collection of sets of the form $\{x \in X_\infty \mid \delta_{x,n} \succ O(\nu)\}$ satisfying (7.8). By the arbitrariness of ε this implies the first statement of Theorem 7.2.

Let s be a solution of equation (7.7). Note that for any measure $\nu \in M_1(\Omega)$, since the logarithm function is concave,

$$\begin{aligned} s \sum_{i=1}^r \nu(i) \ln \theta(i) - \sum_{i=1}^r \nu(i) \ln \frac{\nu(i)}{\mu(i)} &= \sum_{i=1}^r \nu(i) \ln \frac{\mu(i) \theta(i)^s}{\nu(i)} \leq \\ &\leq \ln \left\{ \sum_{\nu(i) > 0} \nu(i) \frac{\mu(i) \theta(i)^s}{\nu(i)} \right\} \leq 0. \end{aligned}$$

Consequently, $d(\nu, \mu, \theta) \leq s$. Now the second part of our theorem follows from the first one if we take $V = M_1(\Omega)$. \square

Remark. In fact the “Bowen equation” is an equation of the form $P(s\varphi) = 0$, where $P(s\varphi)$ is the topological pressure of a weight function $s\varphi$ in a dynamical system (more detailed explanations can be found in [13]). If we replace the topological pressure $P(s\varphi)$ by the spectral potential

$$\lambda(s\varphi, \mu) = \ln \sum_{i=1}^r e^{s\varphi(i)} \mu(i), \quad \text{where } \varphi(i) = \ln \theta(i),$$

then the Bowen equation turns into the equation $\lambda(s\varphi, \mu) = 0$, which is equivalent to (7.7).

8 Block selections of colored brunching processes

Let $\xi_1, \xi_2, \xi_3, \dots$ be a sequence of independent identically distributed random variables taking values 0 or 1 (independent Bernoulli trials).

Lemma 8.1 *If $0 < p' < p < 1$ and $\mathbf{P}\{\xi_i = 1\} \geq p$, then*

$$\mathbf{P}\{\xi_1 + \dots + \xi_k \geq p'k\} \rightarrow 1 \quad \text{as } k \rightarrow \infty \quad (8.1)$$

uniformly with respect to the probability $\mathbf{P}\{\xi_i = 1\} \geq p$.

Proof. In the case $\mathbf{P}\{\xi_i = 1\} = p$ this follows from the law of large numbers. If $\mathbf{P}\{\xi_i = 1\}$ increases then the probability in the left hand side of (8.1) increases as well. \square

Consider a colored brunching process X_1, X_2, \dots with finite set of colors $\Omega = \{1, \dots, r\}$. Each X_n consists of genetic lines (x_1, x_2, \dots, x_n) of length n , in which every subsequent individual has been born by the previous. Fix a (large enough) natural N . We will split genetic lines of length divisible by N into blocks of length N :

$$(x_1, x_2, \dots, x_{nN}) = (y_1, \dots, y_n), \quad \text{where} \quad y_k = (x_{(k-1)N+1}, \dots, x_{kN}).$$

Each block y_k generates an empirical measure δ_{y_k} (spectrum) on Ω by the rule

$$\delta_{y_k}(i) = \frac{\#\{t \mid g(x_t) = i, (k-1)N < t \leq kN\}}{N},$$

where $g(x_t)$ denotes the color of x_t .

A *block selection of order N* from a colored brunching process X_1, X_2, \dots is any sequence of random subsets $Y_n \subset X_{nN}$ with the following property: if $(y_1, \dots, y_{n+1}) \in Y_{n+1}$ then $(y_1, \dots, y_n) \in Y_n$. In this case the sequence of blocks (y_1, \dots, y_{n+1}) will be called a *prolongation* of the sequence (y_1, \dots, y_n) .

As above (see (6.2)), denote by $\mu(i)$ the expectation of children of color i born by each individual, and by μ the corresponding measure on Ω .

Theorem 8.2 *Let X_1, X_2, X_3, \dots be an unconditional colored brunching process with finite set of colors Ω and probability of degeneration $q^* < 1$. If a measure $\nu \in M_1(\Omega)$ satisfies the condition $\rho(\nu, \mu) < 0$, then for any its neighborhood $O(\nu) \subset M_1(\Omega)$ and any number $\varepsilon > 0$ with probability $1 - q^*$ one can extract from the brunching process a block selection Y_1, Y_2, \dots of an order N such that each sequence of blocks $(y_1, \dots, y_n) \in Y_n$ has at least $l(N)$ prolongations in Y_{n+1} , where*

$$l(N) = e^{N(-\rho(\nu, \mu) - \varepsilon)}, \quad (8.2)$$

and the spectra of all blocks belong to $O(\nu)$.

Proof. Fix any numbers p and ε satisfying the conditions

$$0 < p < p + \varepsilon < 1 - q^*, \quad \rho(\nu, \mu) + \varepsilon < 0.$$

By the second part of Theorem 6.3 for all large enough N we have

$$\mathbf{P}\{\#\{x \in X_N \mid \delta_{x,N} \in O(\nu)\} > e^{N(-\rho(\nu, \mu) - \varepsilon/2)}\} \geq p + \varepsilon. \quad (8.3)$$

Further we will consider finite sequences of random sets X_1, \dots, X_{nN} and extract from them block selections Y_1, \dots, Y_n of order N such that the spectra of all their blocks belong to $O(\nu)$ and each sequence of blocks $(y_1, \dots, y_k) \in Y_k$ has at least $l(N)$ prolongations in Y_{k+1} . Denote by A_n the event of existence of a block selection with these properties. Define one more event A by the condition

$$\#\{x \in X_N \mid \delta_{x,N} \in O(\nu)\} > l(N)e^{N\varepsilon/2}.$$

It follows from (8.2) and (8.3) that $\mathbf{P}(A) \geq p + \varepsilon$. Evidently, $A \subset A_1$. Therefore, $\mathbf{P}(A_1) \geq p + \varepsilon$. Now we are going to prove by induction that $\mathbf{P}(A_n) \geq p$ whenever the order N of selection is large enough. Let us perform the step of induction. Assume that $\mathbf{P}(A_n) \geq p$ is valid for some n . Consider the conditional probability $\mathbf{P}(A_{n+1}|A)$.

By the definition of events A_{n+1} and A it cannot be less than the probability of the following event: there are at least $l(N)$ wins in a sequence of $[l(N)e^{N\varepsilon/2}]$ independent Bernoulli trials with probability of win $P(A_n)$ in each. Using Lemma 8.1 (with $p' = p/2$ and $k = [l(N)e^{N\varepsilon/2}]$) one can make this probability greater than $1 - \varepsilon$ at the expense of increasing N . Then,

$$P(A_{n+1}) \geq P(A)P(A_{n+1}|A) > (p + \varepsilon)(1 - \varepsilon) > p.$$

Thus the inequality $P(A_n) > p$ is proved for all n .

It means that with probability greater than p one can extract from the sequence X_1, \dots, X_{nN} a block selection Y_1, \dots, Y_n of order N such that the spectra of all blocks belong to the neighborhood $O(\nu)$ and each sequence of blocks $(y_1, \dots, y_k) \in Y_k$ has at least $l(N)$ prolongations in Y_{k+1} .

To obtain a block selection of infinite length with the same properties, we will construct finite block selections Y_1, \dots, Y_n in the following manner. Initially, suppose that every Y_k , where $k \leq n$, consists of all sequences of blocks $(y_1, \dots, y_k) \in X_{kN}$ such that the spectrum of each block lies in $O(\nu)$. At the first step we exclude from Y_{n-1} all sequences of blocks having less than $l(N)$ prolongations in Y_n , and then exclude from Y_n all prolongations of the sequences that have been excluded from Y_{n-1} . At the second step we exclude from Y_{n-2} all sequences of blocks having after the first step less than $l(N)$ prolongations in the modified Y_{n-1} , and then exclude from Y_{n-1} and Y_n all prolongations of the sequences that have been excluded from Y_{n-2} . Proceeding further in the same manner, after n steps we will obtain a block selection Y_1, \dots, Y_n such that each sequence of blocks from any Y_k has at least $l(N)$ prolongations in Y_{k+1} . Evidently, this selection will be the maximal among all selections of order N having the mentioned property. Therefore with probability at least p all the sets Y_k are nonempty.

For every n let us construct, as is described above, the maximal block selection $Y_1^{(n)}, \dots, Y_n^{(n)}$. From the maximality of these selections it follows that

$$Y_n^{(n)} \supset Y_n^{(n+1)} \supset Y_n^{(n+2)} \supset \dots$$

Define the sets $Y_n = \bigcap_{k \geq n} Y_n^{(k)}$. Then with probability at least p all of them are nonempty and compose an infinite block selection from Theorem 8.2. Since p may be chosen arbitrarily close to $1 - q^*$, such selections do exist with probability $1 - q^*$. \square

Theorem 8.2 can be strengthened by taking several measures in place of a unique measure $\nu \in M_1(\Omega)$.

Theorem 8.3 *Let X_1, X_2, X_3, \dots be an unconditional colored branching process with finite set of colors Ω and probability of degeneration $q^* < 1$. If a finite collection of measures $\nu_i \in M_1(\Omega)$, where $i = 1, \dots, k$, satisfy the inequalities $\rho(\nu_i, \mu) < 0$, then for any neighborhoods $O(\nu_i) \subset M_1(\Omega)$ and any $\varepsilon > 0$ with probability $1 - q^*$ one can extract from the branching process a block selection Y_1, Y_2, \dots of an order N such that for every $i = 1, \dots, k$ each sequence of blocks $(y_1, \dots, y_n) \in Y_n$ has at least*

$$l_i(N) = e^{N(-\rho(\nu_i, \mu) - \varepsilon)}$$

prolongations $(y_1, \dots, y_n, y) \in Y_{n+1}$ with the property $\delta_y \in O(\nu_i)$.

It can be proved in the same manner as the previous one, only now the event A_n should be understood as existence of a finite block selection Y_1, \dots, Y_n satisfying the conclusion of Theorem 8.3 and the event A should be defined by the system of inequalities

$$\#\{x \in X_N \mid \delta_{x,N} \in O(\nu_i)\} > l_i(N) e^{N\varepsilon/2}, \quad i = 1, \dots, k.$$

We leave details to the reader.

9 Dimensions of random fractals (lower bounds)

Now we proceed investigation of the space of infinite genetic lines

$$X_\infty = \{x = (x_1, x_2, x_3, \dots) \mid (x_1, \dots, x_n) \in X_n \ \forall n \in \mathbb{N}\},$$

which is generated by an unconditional colored branching process X_1, X_2, \dots with finite set of colors $\Omega = \{1, \dots, r\}$. It is supposed that there is a measure

$$\mu = (\mu(1), \dots, \mu(r))$$

on Ω , where $\mu(i)$ denotes the expectation of children of color i born by each individual, and X_∞ is equipped with the cylinder θ -metrics (7.1).

Theorem 9.1 *Let X_1, X_2, X_3, \dots be an unconditional colored branching process with finite set of colors Ω and probability of degeneration $q^* < 1$. If a measure $\nu \in M_1(\Omega)$ satisfies the condition $d(\nu, \mu, \theta) > 0$, then with probability $1 - q^*$ for any neighborhood $O(\nu)$ we have the lower bound*

$$\dim_H \{x \in X_\infty \mid \exists N \ \forall n > N \ \delta_{x,n} \in O(\nu)\} \geq d(\nu, \mu, \theta). \quad (9.1)$$

Proof. Fix any number $\alpha < d(\nu, \mu, \theta)$ and so small $\varepsilon > 0$ that

$$d(\nu, \mu, \theta) = \frac{\rho(\nu, \mu)}{\sum_{i=1}^r \nu(i) \ln \theta(i)} > \frac{\rho(\nu, \mu) + 2\varepsilon}{\sum_{i=1}^r \nu(i) \ln \theta(i) - \varepsilon} > \alpha. \quad (9.2)$$

Then choose a convex neighborhood $O^*(\nu)$ whose closure lies in $O(\nu)$ such that for any measure $\delta \in O^*(\nu)$

$$\sum_{i=1}^r \delta(i) \ln \theta(i) > \sum_{i=1}^r \nu(i) \ln \theta(i) - \varepsilon. \quad (9.3)$$

By Theorem 8.2 with probability $1 - q^*$ one can extract from the branching process under consideration a block selection Y_1, Y_2, \dots of order N such that any sequence of blocks $(y_1, \dots, y_n) \in Y_n$ has at least $l(N)$ prolongations in Y_{n+1} , where

$$l(N) = e^{N(-\rho(\nu, \mu) - \varepsilon)},$$

and for each block $y_k = (x_{(k-1)N+1}, \dots, x_{kN})$ the corresponding empirical measure δ_{y_k} (spectrum) belongs to $O^*(\nu)$. Exclude from this selection a certain part of genetic lines in such a way that each of the remaining sequences of blocks $(y_1, \dots, y_n) \in Y_n$ would have exactly $[l(N)]$ prolongations in Y_{n+1} .

Define the random set

$$Y_\infty = \{y = (y_1, y_2, \dots) \mid (y_1, \dots, y_n) \in Y_n, \ n = 1, 2, \dots\}.$$

Any sequence $y = (y_1, y_2, \dots) \in Y_\infty$ consists of blocks of length N . Having written down in order the elements of all these blocks, we obtain from y an infinite genetic line $x = (x_1, x_2, \dots) \in X_\infty$. Denote it as $\pi(y)$. By the definition of Y_∞ the spectrum of each block y_k belongs to $O^*(\nu)$. For every point $x = \pi(y)$, where $y \in Y_\infty$, the empirical measure $\delta_{x,nN}$ is an arithmetical mean of empirical measures corresponding to the first n blocks of y , and so belongs to $O^*(\nu)$ as well. It follows that

$$\pi(Y_\infty) \subset \{x \in X_\infty \mid \exists N \forall n > N \ \delta_{x,n} \in O(\nu)\}. \quad (9.4)$$

The family of all cylinders of the form $Z_{nN}(x)$, where $x \in \pi(Y_\infty)$, generates some σ -algebra on $\pi(Y_\infty)$. Define a probability measure P on this σ -algebra such that

$$P(Z_{nN}(x)) = [l(N)]^{-n}.$$

Then for all large enough N , all $x \in \pi(Y_\infty)$, and all natural n

$$P(Z_{nN}(x)) \leq e^{nN(\rho(\nu, \mu) + 2\varepsilon)}.$$

On the other hand, by (9.3)

$$\begin{aligned} |Z_{nN}(x)| &= \prod_{t=1}^{nN} \theta(x_t) = \exp \left\{ \sum_{t=1}^{nN} \ln \theta(x_t) \right\} = \exp \left\{ \sum_{i=1}^r nN \delta_{x,nN}(i) \ln \theta(i) \right\} \geq \\ &\geq \exp \left\{ nN \left(\sum_{i=1}^r \nu(i) \ln \theta(i) - \varepsilon \right) \right\}. \end{aligned}$$

It follows from the last two formulas and (9.2) that

$$|Z_{nN}(x)|^\alpha \geq \exp \left\{ nN\alpha \left(\sum_{i=1}^r \nu(i) \ln \theta(i) - \varepsilon \right) \right\} \geq e^{nN(\rho(\nu, \mu) + 2\varepsilon)} \geq P(Z_{nN}(x)).$$

Now we are ready to compute the Hausdorff measure of dimension α of the set $\pi(Y_\infty)$. If, while computing the Hausdorff measure, we used coverings of $\pi(Y_\infty)$ not with any cylinders, but with only cylinders of orders divisible by N , then the last formula would imply that such a measure will be at least $P(\pi(Y_\infty)) = 1$. Any cylinder can be put in a cylinder of order divisible by N such that the difference of their orders will be less than N and the ratio of their diameters greater than $\min \theta(i)^N$. Therefore,

$$\text{mes}(\pi(Y_\infty), \alpha) \geq \min \theta(i)^{N\alpha}$$

and hence $\dim_H \pi(Y_\infty) \geq \alpha$.

The set defined in the right hand part of (9.4) contains $\pi(Y_\infty)$. Then its dimension is at least α too. Recall that we have proved this fact by means of a block selection that exists with probability $1 - q^*$. By the arbitrariness of $\alpha < d(\nu, \mu, \theta)$ this implies the desired bound (9.1) with the same probability. \square

Theorem 9.2 *Let s be a root of the Bowen equation*

$$\sum_{i=1}^r \mu(i) \theta(i)^s = 1.$$

If $s \leq 0$, then $X_\infty = \emptyset$ almost surely. Otherwise, if $s > 0$, then X_∞ is nonempty with a positive probability, and with the same probability its dimension equals s .

Proof. The expectation of total number of children of each individual in the brunching process generating the set X_∞ is equal to $m = \mu(1) + \dots + \mu(r)$. If $s \leq 0$, then $m \leq 1$. In this case by Theorem 5.2 our brunching process degenerates almost surely, and $X_\infty = \emptyset$.

If $s > 0$, then $m > 1$. In this case by Theorem 5.2 our brunching process is degenerate with a positive probability, and X_∞ is nonempty. Define a measure $\nu \in M_1(\Omega)$ by means of the equality

$$\nu(i) = \mu(i) \theta(i)^s, \quad i \in \Omega.$$

Then, evidently, $d(\nu, \mu, \theta) = s$. By the previous Theorem $\dim_H X_\infty \geq s$ with the same probability with which $X_\infty \neq \emptyset$. On the other hand, by Theorem 7.2 the inverse inequality holds almost surely. \square

A more general version of Theorem 9.2, in which the similarity coefficients $\theta(1), \dots, \theta(r)$ are random, is proved in [6, 7, 8, 11].

For every probability measure $\nu \in M_1(\Omega)$ define a basin $B(\nu) \subset X_\infty$ as the set of all infinite genetic lines $x = (x_1, x_2, x_3, \dots)$ such that the corresponding sequence of empirical measures $\delta_{x,n}$ converges to ν . What is the dimension of $B(\nu)$? By Theorem 7.2 it does not exceed the Billingsley–Kullback entropy $d(\nu, \mu, \theta)$ with probability 1. On the other hand, the inverse inequality does not follow from the previous results (and, in particular, from Theorem 9.1). To obtain it, we ought to enhance the machinery of block selections.

Lemma 9.3 *Let Q_1, \dots, Q_{2^r} be vertices of a cube in \mathbb{R}^r . Then there exists a choice law $i: \mathbb{R}^r \rightarrow \{1, \dots, 2^r\}$ such that if neighborhoods $O(Q_i)$ are small enough, sequences*

$$\delta_n \in \mathbb{R}^r \quad \text{and} \quad \Delta_n = \frac{\delta_1 + \dots + \delta_n}{n}$$

satisfy the conditions $\delta_{n+1} \in O(Q_{i(\Delta_n)})$ and $\delta_1 \in O(Q_1) \cup \dots \cup O(Q_{2^r})$, then the sequence Δ_n converges to the center of the cube.

Proof. First consider the case $r = 1$, when the cube turns to a segment. Let, for definiteness, $Q_1 = -1$ and $Q_2 = 1$. Set

$$i(\Delta) = \begin{cases} 1, & \Delta \geq 0, \\ 2, & \Delta < 0. \end{cases} \quad (9.5)$$

Take any neighborhoods $O(Q_1)$ and $O(Q_2)$ with radii at most 1. Then for any sequence δ_n satisfying the conditions $\delta_{n+1} \in O(Q_{i(\Delta_n)})$ and $|\delta_1| < 2$ we have the estimate $|\Delta_n| < 2/n$. It may be easily proved by induction. Thus in the one-dimensional case the lemma is proved. To prove it in the multidimensional case one should choose a coordinate system with origin at the center of the cube and axes parallel to edges of the cube and apply the choice law (9.5) to each of the coordinates independently. \square

Theorem 9.4 *Let X_1, X_2, X_3, \dots be an unconditional colored brunching process with finite set of colors Ω and probability of degeneration $q^* < 1$. If a measure $\nu \in M_1(\Omega)$ satisfies the condition $d(\nu, \mu, \theta) > 0$, then with probability $1 - q^*$*

$$\dim_H B(\nu) = d(\nu, \mu, \theta).$$

Proof. Fix any number $\alpha < d(\nu, \mu, \theta)$ and so small $\varepsilon > 0$ that

$$d(\nu, \mu, \theta) = \frac{\rho(\nu, \mu)}{\sum_{i=1}^r \nu(i) \ln \theta(i)} > \frac{\rho(\nu, \mu) + 3\varepsilon}{\sum_{i=1}^r \nu(i) \ln \theta(i) - \varepsilon} > \alpha.$$

The set $M_1(\Omega)$ is in fact a simplex of dimension $r - 1$, where $r = |\Omega|$. Suppose first that ν is an inner point of this simplex (in other words, $\nu(i) > 0$ for all $i \in \Omega$). Take a small convex neighborhood $O(\nu) \subset M_1(\Omega)$ such that for any measure $\delta \in O(\nu)$

$$\begin{aligned} \rho(\delta, \mu) &< \rho(\nu, \mu) + \varepsilon, \\ \sum_{i=1}^r \delta(i) \ln \theta(i) &> \sum_{i=1}^r \nu(i) \ln \theta(i) - \varepsilon. \end{aligned}$$

Let $Q_1, \dots, Q_{2^{r-1}}$ be vertices of some cube in $O(\nu)$ with center at ν . Define for them small neighborhoods $O(Q_i) \subset O(\nu)$ as in Lemma 9.3. By Theorem 8.3 with probability $1 - q^*$ one can extract from the colored branching process X_1, X_2, \dots a block selection Y_1, Y_2, \dots of order N such that for every $i \leq 2^{r-1}$ each sequence of blocks $(y_1, \dots, y_n) \in Y_n$ has at least

$$l(N) = e^{N(-\rho(\nu, \mu) - 2\varepsilon)}$$

prolongations $(y_1, \dots, y_n, y) \in Y_{n+1}$ possessing the property $\delta_y \in O(Q_i)$.

Exclude from this block selection a certain part of genetic lines so that each of the remaining sequences of blocks $(y_1, \dots, y_n) \in Y_n$ would have exactly $[l(N)]$ prolongations $(y_1, \dots, y_n, y) \in Y_{n+1}$, and all these prolongations would satisfy the choice law from Lemma 9.3, namely,

$$\delta_y \in O(Q_{i(\Delta_n)}), \quad \text{where} \quad \Delta_n = \frac{\delta_{y_1} + \dots + \delta_{y_n}}{n}.$$

Denote by $\pi(Y_\infty)$ the set of all infinite genetic lines $(x_1, x_2, \dots) \in X_\infty$ for which every initial segment of length nN , has been partitioned into blocks of length N , turns into an element of Y_n . Then by Lemma 9.3 we have the inclusion $\pi(Y_\infty) \subset B(\nu)$.

Reproducing reasoning from the proof of Theorem 9.1 one can ascertain that the dimension of $\pi(Y_\infty)$ is greater than α . Since α can be taken arbitrarily close to $d(\nu, \mu, \theta)$, we obtain the lower bound $\dim_H B(\nu) \geq d(\nu, \mu, \theta)$. The inverse inequality, as was mentioned above, follows from Theorem 7.2. Thus in the case of inner point $\nu \in M_1(\Omega)$ the theorem is proved.

If the measure ν belongs to the boundary of the simplex $M_1(\Omega)$, then one should exclude from Ω all elements i with $\nu(i) = 0$, and consider the set

$$\Omega' = \{i \in \Omega \mid \nu(i) > 0\}.$$

Exclude from the brunching process X_1, X_2, \dots all genetic lines containing elements of colors not in Ω' and denote as X'_1, X'_2, \dots the resulting brunching process (with the set of colors Ω'). The corresponding set of infinite genetic lines X'_∞ is contained in X_∞ . It follows from the definition of Billingsley–Kullback entropy $d(\nu, \mu, \theta)$ that it is the same for the sets of colors Ω and Ω' . Besides, the measure ν lies in the interior of the simplex $M_1(\Omega')$. Therefore, $\dim_H B(\nu) \cap X'_\infty = d(\nu, \mu, \theta)$ with the same probability as $X'_\infty \neq \emptyset$.

The theorem would be completely proved if the probability of the event $X'_\infty \neq \emptyset$ was equal to $1 - q^*$. But it may be less than $1 - q^*$. This obstacle can be overcome as follows.

Let $m' = \sum_{i \in \Omega'} \mu(i)$. This is nothing more than the expectation of each individual's number of children in the brunching process X'_1, X'_2, \dots . If $m' \leq 1$ then (2.4) implies the inequality $\rho(\nu, \mu) \geq 0$, which contradicts the condition $d(\nu, \mu, \theta) > 0$ of our theorem. Therefore $m' > 1$ and, respectively, the probability of the event $X'_\infty = \emptyset$ is strictly less than 1. Let us denote it q' .

If the brunching process X'_1, X'_2, \dots was generated not by a unique initial element, but k initial elements, then the probability of $X'_\infty = \emptyset$ would be equal to $(q')^k$. Recall that the cardinality of X_n grows exponentially with probability $1 - q^*$. If this is the case, one can first wait for the event $|X_n| \geq k$, and then consider separately $|X_n|$ independent counterparts of the brunching process X'_1, X'_2, \dots generated by different elements of X_n . This trick allows to obtain the bound $\dim_H B(\nu) \geq d(\nu, \mu, \theta)$ with conditional probability at least $1 - (q')^k$ under the condition $|X_n| \rightarrow \infty$. Since k is arbitrary, the above mentioned conditional probability is in fact one, and the complete probability cannot be less than $1 - q^*$. \square

References

- [1] *Athreya K. B., Ney P. E.* Branching processes. Springer-Verlag, 1972. 287 p.
- [2] *Billingsley P.* Hausdorff dimension in probability theory // *Ill. J. Math.* **4** (1960), 187–209.
- [3] *Billingsley P.* Hausdorff dimension in probability theory. II // *Ill. J. Math.* **5** (1961), 291–298.
- [4] *Deuschel J.-D., Stroock D.* Large deviations. Pure and applied mathematics, 137. Boston: Academic Press, 1989. 307 p.
- [5] *Eggleston H. G.* The fractional dimension of a set defined by decimal properties // *Q. J. Math., Oxf. Ser.* **20** (1949), 31–36.
- [6] *Falconer K. J.* Random fractals // *Math. Proc. Camb. Philos. Soc.* **100** (1986), 559–582.
- [7] *Falconer K.* Fractal geometry. Mathematical foundations and applications. 2-nd ed. Chichester: Wiley, 2003. 337 p.
- [8] *Graf S.* Statistically self-similar fractals // *Probab. Theory Relat. Fields* **74** (1987), 357–392.

- [9] *Harris T. E.* The theory of branching processes. Springer-Verlag, 1963. 230 p.
- [10] *Hutchinson J. E.* Fractals and self-similarity // Indiana Univ. Math. J. **30** (1981), 713–747.
- [11] *Mauldin R. D., Williams S. C.* Random recursive constructions: asymptotic geometric and topological properties // Trans. Am. Math. Soc. **295** (1986), 325–346.
- [12] *Moran P. A. P.* Additive functions of intervals and Hausdorff measure // Proc. Camb. Phil. Soc. **42** (1946), 15–23.
- [13] *Pesin Ya. B.* Dimension theory in dynamical systems: contemporary views and applications. Chicago University Press, 1998. 304 p.
- [14] *Varadhan S. R. S.* Large deviations // Ann. Probab. **36**, No. 2 (2008), 397–419.
- [15] *Young L.-S.* Dimension, entropy and Lyapunov exponents // Ergod. Theory and Dyn. Syst. **2** (1982), 109–124.